科学数据的控制策略及其模型优化*

文禹衡[1,2]　徐坤杉[1]
[1]湘潭大学知识产权学院　湘潭　411105
[2]湖南省数据治理与智慧司法研究中心　湘潭　411105

摘　要：［目的/意义］提出并优化科学数据的控制策略及其模型，有助于推进科学数据平台的基础制度建设，平衡科学数据安全保护与共享利用。［方法/过程］选择国家级科学数据平台的用户协议等为样本，运用扎根理论方法获取 76 个初始概念、9 个副范畴和 5 个主范畴，进而提炼科学数据的控制策略模型。［结果/结论］我国科学数据平台已形成以目的、行为、权利、身份和责任为构成要素的“五控一体”控制策略模型，建议从事前防御、事中管理和事后归责方面，优化形成“三阶段”控制策略模型。具体措施包括：①概括列举目的类型，强化目的控制；②明确数据权利归属，加强权利控制；③统一分级分类管理，补强身份控制；④平衡平台管理与用户控制，转型平台自治，促进平台意识与规范统一，新增行业他律；⑤合理分配平台、用户和科学数据提供者的责任。

关键词：科学数据　控制策略　扎根理论　行为控制　权利控制
分类号：G203

1 引言

2022 年，《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》（以下简称《数据二十条》）指出数据基础制度建设事关国家发展和安全大局[1]。2023 年，国家数据局的组建意味着我国数据基础制度建设进入新阶段。科学数据是国家战略性资源之一，与科学数据平台（或中心）相关的基础制度建设同样应备受重视。早在 2018 年，《科学数据管理办法》（以下简称《办法》）就对科学数据进行了界定。《办法》规定科学数据应以开放为常态、不开放为例外为原则[2]，确立了我国科学数据平台建设应当以开放、共享为价值导向。在此背景下，科学数据平台不仅需要合理地管理已有科学数据，筑牢科学数据安全屏障，还应当安全有序地开展科学数据共享，一套合理有效的科学数据控制策略至关重要。形成切实可行、系统科学的控制体系，方能建设优质的科学数据平台，满足国内优势产业的发展需求，起到支撑国家科技创新、经济社会发展和国家安全的作用。

截至 2023 年 3 月，进行预检时发现，国内外关于科学数据的研究成果非常丰富。例如，国内成果就涉及科学数据的元数据、利益相关者、质量、素养、管理、平台（中心）建设、

*本文系国家社会科学基金青年项目“数据要素确权的法律供给研究”（项目编号：21CFX007）研究成果之一。
作者简介：文禹衡，湖南省数据治理与智慧司法研究中心主任，副教授，博士，硕士生导师，E-mail:jsjmyh@163.com；徐坤杉，硕士研究生。

开放共享、安全保障、服务、引用、关联、融合、使用、出版、监管、隐私保护以及与个人数据保护协同等方面，各方面还可以继续细分主题。为了精简篇幅和集中主题，在确定现有成果未涉及“科学数据的控制策略”之后，考虑到本文的样本来源于科学数据平台（中心），故将文献综述内容聚焦到科学数据平台（中心）。

在中国知网学术期刊（CSSCI），以“科学数据平台”“科学数据中心”“科学数据共享平台”和“科学数据共享中心”为检索词进行篇名精确检索。我国目前以科学数据（共享）平台/中心作为研究对象的研究成果相对较少，主要涉及：①实践状况，例如“互联网+”环境有助于形成科学数据共享的深度融合、人人参与的优势环境，为完善科学数据共享平台的构建和保障科学数据资源的高效利用提供了有利条件[3]，国内现有平台在组织管理、资源建设、用户服务多方面探索力求形成相对完整的基础建设[4]，但仍存在服务方式单一、强化数据主导而忽视数据服务[5]、规范性说明不足、内容整合层次较浅、资源获取不够直接[6]等问题。②专业性平台，涉及国家农业科学数据共享中心[7]、中国西部环境与生态科学数据中心[8]、开放医学科学数据平台[9]、黄河流域专题数据的国家科学数据中心[10]、长三角跨区域科学数据中心[11]、野外科学观测研究台站（网络）和科学数据中心[12]、复旦大学社会科学数据平台[13]，以及湖南省科技信息与科学数据共享平台[14]等多领域的科学数据中心。③具体建设内容，主要分布在科学数据共享平台的绩效评估[15]与服务效能分析[16]、引用规范调查研究[17]、被引用状况调查分析[18]、数据管理研究[19]、用户注册协议合规性[20]等方面，也有学者聚焦于开源技术[21]、知识管理[22]等应用技术方面。④国外数据库研究启示，有学者调查了德国地球科学领域科学数据中心，提出符合我国相关领域数据中心的发展概念图[23]；还有学者研究欧美生物医学科学数据中心后，从基础保障、基础设施、资源建设、技术研发、标准规范、用户服务 6 个方面提出启示[24]。

在 Semantic scholar 检索 2014 年以来的英文文献，以“scientific data platform”“scientific data center”“scientific sharing data platform”和“scientific sharing data center”进行篇名检索，并以 Google scholar、Web of science 为辅助的数据来源进行补充检索，人工筛选前述时间范围内的文献。近 10 年来，国外相关针对科学数据（共享）平台（中心）的研究主要涉及：①实践状况，例如开放科学数据中心的实践现状[25]、科学数据中心的数据管理原则[26]；②专业性平台，例如开放的电子科学数据共享平台（B2SHARE）[27]、虚拟分子科学数据中心（VHERLA）[28]等；③具体建设内容，例如科学数据的应用系统[29]、是否云端化[30]、目录管理[31]等。

综上所述，国内外研究都侧重于科学数据平台建设，在实践状况、专业平台建设与具体建设内容方面形成了一定的成果，国内还注重国外数据库研究启示，对科学数据控制策略关注不够。鉴于此，本文选择国家级科学数据中心为研究对象，分析平台实践中的科学数据控制策略，提炼模型并提出相应优化建议，以期引发学界对“控制策略”主题的关注，进而推动形成科学数据控制策略理论。

2 研究过程

2.1 筛选样本

基于国家科技基础条件平台中心，获取 20 个国家科学数据中心和 31 个国家资源库名录（以下简称“国家级科学数据平台”）。在各平台官网检索用户注册协议、科学数据共享政策或用户使用说明等可能涉及科学数据管理内容的相关文件（以下统称“用户协议”），但大部分平台并没有相关文件，用户通过注册可直接登录网站获取数据，有 9 个平台无需注册即可下载数据，最后获取 11 个国家科学数据中心和 3 个国家资源库的用户协议。通过人工阅读协议全文，共提取 51 个涉及科学数据的原始语段（见表 1）。需要注意的是，由于多数平台笼统地将服务条款、免责声明、共享政策等置于一个文件中，不同平台的文件不一致，原始语段在各样本中的位置也不一样，既可能在隐私政策部分，也可能在免责声明等部分。

表 1 涉及科学数据内容的国家级科学数据平台样本

Table 1 Samples of national scientific data platform including scientific data content

样本编号	平台名称	原始语段所在位置	网址	原始语段数 /个
1S	国家微生物科学数据中心	数据使用说明	https://nmdc.cn/	1
2S	国家基因组科学数据中心	隐私政策；免责声明	https://ngdc.cncb.ac.cn/	1
3S	国家青藏高原科学数据中心	隐私政策；使用条款；免责声明	https://data.tpdc.ac.cn/	5
4S	国家计量科学数据中心	用户注册协议	https://www.nmdc.ac.cn/	4
5S	国家地球系统科学数据中心	用户服务协议	http://www.geodata.cn/	3
6S	国家人口健康科学数据中心	隐私和安全政策；免责声明；用户注册须知	https://www.ncmi.cn/phda/support.html?type=md028	4
7S	国家基础学科科学数据中心	用户服务协议；免责声明	https://nbsdc.cn/	2
8S	国家林业和草原科学数据中心	服务说明；数据共享政策；数据安全保密协议书	http://www.forestdata.cn/	5
9S	国家气象科学数据中心	服务条款；用户权益	http://data.cma.cn/	4
10S	国家地震科学数据中心	用户注册协议	https://data.earthquake.cn/index.html	2
11S	国家海洋科学数据中心	用户许可协议	http://mds.nmdis.org.cn/	3
12S	国家海洋水产种质资源库	共享服务管理办法	http://marine.fishinfo.cn/	14
13S	国家重要野生植物种质资源库	隐私政策、服务条款	https://seed.iflora.cn/	2
14S	国家植物标本资源库（NPSRC）	注册用户使用协议	https://www.cvh.ac.cn/	1

2.2 选择方法

在质性研究领域，格拉斯（B. Glaser）和斯特劳斯（A. Strauss）提出的“扎根理论”（grounded theory）被广泛运用。国内学者考究认为，扎根理论的主要宗旨是从经验资料的基础上建立理论——研究者在研究开始之前一般没有理论假设，直接从实际观察入手，从原始资料中归纳出经验概括，然后上升到理论[32]。不同于运用扎根理论方法辄止于构建理论[33]或模型[34]、归纳因素[35]等，即编码后就落脚到理论、模型或因素之上，本文的研究目标有两个：①尽可能客观地呈现科学数据控制策略的实然现状，即“实然模型”是什么样的；②在此基础上提出科学数据控制策略的优化方案，即“应然模型”是怎么样的。为了实现这两

个目标，主要运用扎根理论方法，辅之以词频统计方法。

呈现实然现状的传统做法是依靠假设、经验和归纳，基本上是从素材直接到结果，往往会因为研究者的认识、理解和价值取向等因素，导致结果存在较大误差。扎根理论方法一开始就不做任何预设，从初始概念就“忠实”于原始素材，再逐步完成范畴、副范畴或主范畴，编码后进行饱和度检验，能尽可能确保结果的客观性和全面性。此外，从收集整理的样本来看，科学数据的相关内容分布在不同平台的用户协议，涉及用户注册、科学数据共享或使用说明等相关内容，而扎根理论方法适合规模大、碎片化样本的定量和定性分析[36]。扎根理论在发展过程中形成了经典扎根理论（开放编码、选择编码、理论编码）、程序化扎根理论（开放编码、主轴编码、选择编码）和构建主义扎根理论（初级编码、聚焦编码、轴心编码和理论编码）3 种不同范式，当前程序化扎根理论方法仍是我国图情研究应用的主流[37]。考虑到归纳科学数据控制策略的内容需要开放编码、主轴编码、选择编码，故具体选择程序化扎根理论方法。

诚然，运用扎根理论方法与量化实证研究不同，研究者在进入田野调查之前并不提出理论假设，而是直接从调查资料中进行经验概括，最终提升为理论[38]。如果只是为了提炼科学数据管理策略理论模型，不需要统计概念、范畴、主范畴的频次。本文之所以在此基础上进行词频统计，是试图揭示科学数据控制策略的“实然模型”的各构成要素的比重，由此量化要素在实践中被重视的程度。因本文是基于全样本展开分析，这些样本反映的是科学数据控制策略的整体现状，所以基于概念、范畴、主范畴的频次和占比能够客观反映其所指代对象在整体现状中的重要性。由于各样本来自不同国家级科学数据平台，每个平台对于科学数据的规定，既有共性之处，又有个性之处，统计频次并计算占比具有一定意义。在总体样本数量一定时，占比能够说明一些问题——占比小不一定是个性之处，但占比大在较大程度上说明具有共性之处。虽然不能仅依靠占比来说明其重要性，但是占比确实能够反映实践中的通用做法。例如，行为控制出现 71 次、占比 69.61%，说明行为控制是科学数据控制策略的通用实践；目的控制出现 7 次、占比 6.86%，说明目的控制在科学数据控制策略中较少被运用。

鉴于此，运用词频统计方法有助于明确科学数据控制策略的内部要素及其组合的优化思路。①如果某要素在“实然模型”中比重高：根据专家经验认为其对于科学数据控制策略重要，则应建议维持或提升；根据专家经验认为其对于科学数据控制策略不重要，则应建议降低。②如果某要素在“实然模型”中比重低：根据专家经验认为其对于科学数据控制策略很重要，则应建议适当提升；根据专家经验认为其对于科学数据控制策略可有可无，则应建议维持或降低。总的来说，本文试图在运用扎根理论方法进行质性分析的基础上，引入词频作为辅助方法完成定量分析，尽可能使结论可靠、建议合理。

2.3 处理样本

2.3.1 开放性编码

开放性编码是从原始语段中提取重点语句，形成初始概念的过程，提取重点语句有助于使初始概念更加“忠实”于原始素材。编码过程中，一个原始语段往往蕴含多个重点语句，原始语段与初始概念基本上是一对多，很少是一对一。鉴于篇幅，考虑到样本 3S 涵盖“一对一”“一对多”情形且篇幅适中，故以其作为开放性编码示例（见表 2）。为了呈现提炼的思维过程，在表 2 中呈现了重点语句提取步骤。

表 2 科学数据原始语段的开放性编码（示例）

Table 2 Open coding of scientific data primitives (examples)

样本编号	原始语段内容	原始语段编号	重点语句提取	初始概念
3S	“国家青藏高原科学数据中心”网站上的一切数据和资料版权归数据和资料原生产单位所有	3S01001	网站中一切数据和资料版权归原生产单位所有	版权归原生产单位
	为保障数据开发者的著作权，用户在使用全部或部分“国家青藏高原科学数据中心”所产出的研究成果中须在相关成果的显著位置上明确注明数据来源	3S02001	须在相关成果显著位置明确注明数据来源	使用须标明出处
	不得将数据转让给第三方，由此引起的一切后果由数据用户承担	3S03001	不得将数据转让给第三方	不得转让
		3S03002	一切后果由数据用户承担	用户自担风险
	从数据中心获取的数据如属于数据保密范围，用户须与“国家青藏高原科学数据中心”签署保密协议，如出现问题，用户承担完全责任	3S04001	使用保密数据须签署保密协议	保密数据使用要求
		3S04002	如出现问题，用户承担完全责任	用户自担风险
	下载（包括印出）的任何材料都含有版权提示信息，用户不得删改相关版权提示信息，防止该文件的全部或部分被用于其它目的；删改版权提示信息的用户将据著作权法承担版权侵权责任	3S05001	用户不得删改相关版权的提示信息	不得删改版权提示信息
		3S05002	用户删减版权提示信息应承担版权侵权责任	承担版权侵权责任

2.3.2 主轴性编码

主轴性编码是在初始概念的基础上提取副范畴，并合并同类副范畴得到主范畴的过程。一个初始概念可提取一个副范畴，但是在汇总各初始概念和副范畴时，前者与后者是多对一的数量关系。同样，副范畴和主范畴也是多对一的数量关系。如表 3 所示，在 76 个初始概念的基础上进行提炼，共获取 9 个副范畴，即使用行为要求、获取行为要求、管理行为要求、使用目的限制、风险责任承担、获取身份限制、知识产权归属、数据权利归属和用户权利限制；在此基础上获取 5 个主范畴，即行为控制、目的控制、责任控制、身份控制和权利控制。

表 3 科学数据控制策略主轴性编码汇总

Table 3 Spindle coding of scientific data control strategy

初始概念（频次 /次）	副范畴（频次 /次）	主范畴（频次 /次）	初始概念（频次 /次）	副范畴（频次 /次）	主范畴（频次 /次）
1-4 级用户可互相使用数据	使用行为要求（46）	行为控制（51）	违反规范平台将采取措施	管理行为要求（20）	行为控制（20）
保护机密			数据编号管理		
保密数据使用受限			数据保护期依用户级别确定		
不得进行转售或商业利用（3）			设立专门机构		
不得删改版权提示信息			安排专人负责		
不得衍生利用			按平台规定录入数据并报送		
不得转让（2）			提交 II 类数据的附加条件		
超出保护期限的数据可免费申请使用			每月另存、每年刻录存档		

优先使用网站提供的引用方式和标识			数据维护更新需专用服务器和固定宽带网络连接		
仅用于已批准研究			数据维护更新需专门人员		
禁止出售（2）			备份需专用措施、专人定期		
禁止识别（2）			拟更新数据需专人审核		
禁止违规共享			确认更新的数据需提交审查		
访问受控数据需获批			需专人负责数据更新维护		
使用需包括版权声明			本地数据经审核可远程更新		
使用须标明出处（6）			通过审核的数据可公开发布		
中英使用均需注明原始来源			成员单位可独立提供数据服务		
使用须保留作者或授权人对原作品的署名			平台负责发布数据、审批申请、提供数据服务		
使用需获得权利人许可			平台负责对外发布并说明		
使用需经授权			平台提供服务配备专门设备		
使用需寄送成果（2）			不得用于商业目的（5）	使用目的的限制（7）	目的控制（7）
受控数据使用规范			需以获得信息为目的		
数据生产者自由使用或处置所生产的数据			严禁任何未经本协议明确允许的目的使用		
未经明确书面同意不许下载或修改平台或其任何部分			知识产权归本平台（4）	知识产权归属（5）	权利控制（9）
未经书面同意不得发布本平台任何内容			版权归原作者与平台共有		
未经事先书面同意不得以任何方式侵犯知识产权			原始数据和数据信息的权限归原单位所有（2）	数据权利归属（2）	
未经书面许可不得对外提供（3）			用户享有有限使用权（2）	用户权利限制（2）	
未经书面许可不得修改或销售			公开数据不限身份均可获取	获取身份限制（9）	身份控制（9）
未经书面许可严禁创建或编辑文集、汇编、数据库或人名地址录			相关部门经认可后可免费获取数据		
无书面许可不得修改、传播及销售			协议共享数据限协议用户获取		
未经许可不得擅自使用			协议公开数据限协议用户获取		
未经允许不得用于外部使用			高级用户获取数据限制		
需提交用途与人员信息			普通用户获取数据限制		
协议用户按流程获取协议共享数据	获取行为要求（5）		公开数据不限身份自由获取		
保密数据申请批准后获取			注册用户权限		
用户可向成员单位申请获取数据			非实名注册用户受限		
用户可向平台申请获取数据			用户自担风险（5）	风险责任承担（6）	责任控制（6）
保护期外数据经申请获取			承担版权侵权责任		

注：该表括号中的数字是表示频次，若没有括号和数字则表示频次为 1

2.3.3 选择性编码

分析副范畴之间的关系可发现，获取身份限制、使用目的限制、风险责任承担、用户权利限制、知识产权归属和数据权利归属均对使用行为要求、获取行为要求或管理行为要求产生影响，实质上仍是通过影响用户或平台的行为实现控制效果（见表 4）。在此基础上，分析主范畴之间的关系（见图 1）：行为控制在科学数据平台实施科学数据控制策略时发挥核心作用，目的、责任、身份、权利的控制也会通过影响行为而起到控制效果。

表 4 科学数据控制策略选择性编码

Table 4 Selective coding of scientific data control strategy

范畴关系	代表性重点语句	关系内涵解读
目的控制→行为控制	所有的资料和图象均不得用于商业目的	通过限制用户使用目的，禁止用户的商用行为
责任控制→行为控制	您对此类数据的任何接收、依赖或使用均由您自行承担风险。NGDC 不对相	通过事前告知责任承担，提醒

	关信息的错误、不准确或遗漏承担任何法律责任	用户谨慎使用科学数据
身份控制→行为控制	协议用户可直接下载协议公开数据;普通用户需到个人中心页面完善个人资料，申请升级为协议用户，待管理员审核通过，方可下载获取协议公开数据	设置平台审核来限制用户身份资格，限制用户获取数据
权利控制→行为控制	网站中一切数据和资料版权归原生产单位所有	通过事先声明权利归属，提醒用户仅可有限使用科学数据

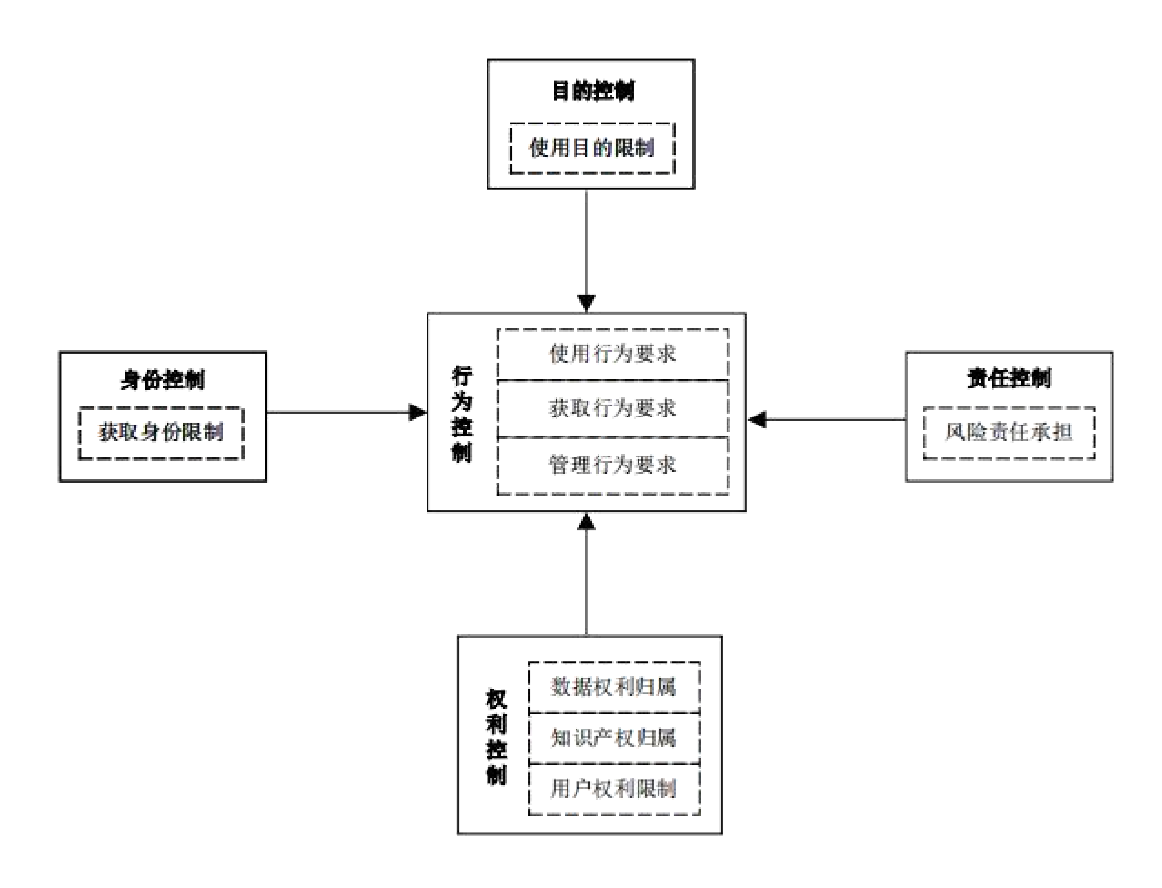


图 1 科学数据的“五控一体”控制策略模型

Figure 1 Model of the “five-control-in-one” strategy of scientific data

2.3.4 饱和度检验

在运用程序化扎根理论编码时：如果是抽样编码，则理论饱和度检验必不可少；如果是全样本编码，可以选择检验饱和度[39]。本文需掌握科学数据控制策略现状，即分析模型内部构成要素（以副范畴、主范畴为表征）的占比，故需要完成全样本编码，在此意义上可视为已达到饱和。但本文仍然进行饱和度检验：在编码到第 10 个样本时，副范畴和主范畴均不再出现新类型，继续完成第 11—13 个样本文件编码，也没有发现新的副范畴和主范畴；此外，将随机预留的编号 14S 的样本作为理论饱和度检验之用，未将其与其他样本一起编码，而是在其他样本编码全部完成后再对其进行编码，其编码结果仍与之前得到的副范畴、主范畴相符（见表 5）。通过前述双重验证，可认为本文编码达到了饱和。

表 5 科学数据控制策略编码的饱和度检验

Table 5 Coding saturation test of scientific data control strategy

编号	平台名称	原始语段	编码	重点语句提取	初始概念	副范畴	主范畴
14S	国家植物标本资源库（NPSRC）	查看完整植物标本采集地（非实名注册用户只能查看采集省份）	14S01001	注册用户可以查看完整的植物标本采集地	注册用户权限	获取身份限制	身份控制
			14S01002	非实名注册用户只能查看植物标本的采集省份	非实名注册用户受限	获取身份限制	身份控制

3 研究结论

3.1 三种行为策略涉及两类主体

行为控制是指通过规定或限制主体行为实现控制的策略，涉及两类主体：①用户，通过设置限定条件来限制用户的使用行为或获取行为，此时的控制策略概括为“使用行为要求”和“获取行为要求”；②平台，详列出本平台管理科学数据的措施，力图实现对科学数据的有序管理，其背后体现了对已有科学数据的正向控制理念，此时的控制策略概括为“管理行为要求”。

3.1.1 用户使用行为控制：选择空间非常之小

“使用行为要求”出现 46 次，在行为控制策略中占比 64.79%，说明使用行为要求是控制科学数据的通用实践，也即控制策略中的重点所在。对“使用行为要求”进行二次分类发现，使用行为要求可分禁为、应为和可为 3 种，其中：禁为规定频次为 24 次，在使用行为要求中占比超过半数；应为规定频次为 18 次；可为规定频次最少（见表 6）。这说明在使用行为要求方面，平台为用户划分了较为清晰的边界，带有强制意味的“禁为”和“应为”占比 91.30%，说明用户“可为”的选择空间非常之小。

表 6 “使用行为要求”的二次分类

Table 6 Secondary classification of “use behavior requirements”

<table>
<tr><th>初始概念（频次 /次）</th><th>性质（频次 /次）</th><th>初始概念（频次 /次）</th><th>性质（频次 /次）</th><th>初始概念（频次 /次）</th><th>性质（频次 /次）</th></tr>
<tr><td>不得转让（2）</td><td rowspan="8">禁为规定（24）</td><td>使用需经授权（2）</td><td rowspan="8">应为规定（18）</td><td rowspan="2">超出保护期限的数据免费申请即可使用（1）</td><td rowspan="8">可为规定（4）</td></tr>
<tr><td>不得删改版权提示信息（1）</td><td>使用需标明出处（7）</td></tr>
<tr><td>不得衍生利用（1）</td><td>使用需包括版权声明（2）</td><td rowspan="2">数据生产者自由使用或处置所生产的数据（1）</td></tr>
<tr><td>不得进行商业利用（3）</td><td>使用需将成果寄送存档（2）</td></tr>
<tr><td>未经许可不得擅自使用（8）</td><td>仅用于已批准研究（1）</td><td rowspan="2">1—4 级用户可以互相使用数据（1）</td></tr>
<tr><td>禁止违规共享(5)</td><td>保护机密（1）</td></tr>
<tr><td>禁止识别（2）</td><td>受控数据使用规范（2）</td><td rowspan="2">请优先使用网站提供的引用方式和标识（1）</td></tr>
<tr><td>禁止出售（2）</td><td>需提交用途与人员信息</td></tr>
</table>

3.1.2 用户获取行为控制：通过申请程序实现

“获取行为要求”仅出现 5 次，在行为控制策略中占比 7.04%，说明获取行为在控制科学数据的实践中很少被运用。原因可能在于获取只是使用数据的步骤之一，用户的意图多在于使用，多数平台出于节省成本或避免重复规定等考量，侧重对使用行为作出要求，而没有过多关注获取行为。在现有规定中，平台的科学数据控制思路体现在前置申请程序设置上，例如“协议用户按流程获取协议共享数据”是依据不同用户等级而设置，“保密数据申请批

准后获取”和“保护期外数据经申请获取”是依据不同数据分级而设置，“用户可向成员单位申请获取数据”和“用户可向平台申请获取数据”是依据不同提供主体而设置。虽然“依据”各有不同，但都是针对获取过程中的特殊情况规定“拟获取，先申请”，若用户没有事先申请，则无法直接从平台处获取到相应的科学数据。

3.1.3 平台管理行为控制：形成较完备的体系

“管理行为要求”共出现 20 次，在行为控制策略中占比 28.17%，说明管理行为在控制科学数据中有所实践，但被运用得不够充分。尽管如此，管理行为已形成了较完备的体系：编号管理、专人专任、配备专门设备、定期审查、是否公布、如何公布、是否更新、如何更新、如何保存等管理措施，贯通数据的收集、传输、存储、处理和交换各阶段，而且部分规定较为细致，便于平台的日常管理工作。

3.2 四类其他策略涵盖多元内容

3.2.1 目的控制：以是否具有商业性为判断

目的控制以用途是否具有商业性为标准，平台许可具有非商业目的用户获取、使用科学数据，排除商业性用途。目的控制出现 7 次，在全部样本中占比 6.86%，呈现 3 种不同模式：①出现最多的是“反向”规定，即“不得用于商业目的”，不明示用户需要具有何种目的，而仅禁止商业目的；②“正向”规定出现 1 次，即“需以获得信息为目的”，此处宜适用限缩解释，即用户仅以“获知”为目的；③“概括”规定出现 1 次，即“严禁任何未经本协议明确允许的目的使用”，平台未给出明确允许的目的说明或列表。总体来看，反向规定和正向规定中的目的都是具体明确的，前者给予用户的使用空间最大，后者给予用户的使用空间最小，而概括规定是模糊的，用户使用空间可大可小，不便于用户判断。

3.2.2 权利控制：产权归属与用户权利限制

权利控制包括两方面内容：①平台所提供科学数据的知识产权或数据权利归属；②用户对已获取的科学数据享有何种权利。权利控制出现 9 次，在全部样本中占比 8.82%，其中知识产权归属出现 5 次、数据权利归属出现 2 次，分属于不同平台。从主体上看，涉及平台方、原单位或原作者三方主体，各平台对知识产权、数据权利的归属尚未达成共识，既有“归平台所有”，也有“归原单位所有”，还有“归原作者与平台共有”。从依据上看，依据来源或是平台投入来判断归属，是平台的“一家之言”，有待进一步明确权利归属的正当性依据。此外，用户权利限制出现 2 次，意味着赋予用户的权利很少——实际上仅赋予了使用权，且进行严格的限制，例如用户对提供的地震科学数据、气象资料只享有有限的、不排他的使用权。限制条件“有限的”“不排他”的积极意义在于，防止科学数据被滥用或垄断。

3.2.3 身份控制：资格等级决定可获取数据

身份控制是指在用户获取科学数据时需要具有的身份资格等级，譬如注册用户、实名认证用户、协议用户等。身份控制出现 9 次，在全部样本中占比 8.82%，内容上均表意为获取

身份限制，即相应身份资格的用户仅能获取相对应等级的科学数据。此处所说“相应等级”并非严格遵循某项标准，而是平台自行规定的数据分级分类制度，对非本平台的科学数据并不适用。对于用户来说，获取相应的身份资格需要完成二次申请，如“协议公开\共享数据仅限协议用户获取”中的“协议用户”，需要使用者在成为普通用户后再次提交申请，待平台进一步审核通过后才能成为协议用户。本质上，身份控制是通过多次审核实现加强平台控制力的目的。

3.2.4 责任控制：平台与用户责任分配失衡

责任控制旨在确定出现平台或用户不利后果时，应该由哪一方来承担责任。责任控制出现 6 次，在全部样本中占比 5.88%，可以归纳为两类：①整体性规定，诸如“用户自担风险”，意指一切风险由用户自行承担；②具体性规定，诸如“承担版权侵权责任”，是明确告知用户未经许可删减版权信息的行为后果。进一步分析可知，平台设置整体性规定可能旨在减轻自身的审查义务，例如“用户接收、依赖或使用平台所提供数据均自行承担风险，NGDC 不对相关信息的错误、不准确或遗漏承担任何法律责任”。考虑到科学数据是由该领域内的科学机构或学者提供，科学数据平台作为媒介，进行专业性的实质审查的能力和成本有限，列出此类规定也在情理之中。倘若能将此类整体性规定向具体性规定转化，分不同情形明确责任主体，则有助于各主体采取积极行为防止不利后果发生。

4 优化建议

科学数据控制策略的“五控一体”模型（见图 1）是基于全样本，通过扎根理论方法提取行为、权利、身份、目的和责任 5 个要素而构建的“实然模型”。该模型体现了科学数据控制策略的现实情况，从应然角度审视，其还存在优化的空间。以“事前—事中—事后”三阶段控制理论审视“五控一体”模型时，发现这 5 个要素可以分置在不同的控制阶段：①权利控制、身份控制与目的控制侧重于在行为发生前约束主体，归于事前控制；②行为控制一般针对行为本身进行规定，归于事中控制；③责任控制主要明确行为发生后的不利后果承担，归于事后控制。鉴于此，优化科学数据控制策略，可以考虑从“五控一体”模型转向“三阶段”模型（见图 2），实现事前—事中—事后三阶段控制。“三阶段”模型仍以“五控一体”模型为基础，并未改变其“目的”“身份”“权利”“行为”和“责任”这 5 个内部构成要素，只不过调整了除“行为”以外要素的位置关系，并就调整位置后的各要素相应提出内容上的优化。

将以“行为”为中心的“目的”“身份”“权利”和“责任”散点分布，调整为“目的、身份和权利归入事前阶段”“行为归入事中阶段”“责任归入事后阶段”的三阶段分布。各阶段的数位命名（前、中、后）并不代表三阶段是单向的线性关系，行为控制的核心地位并未改变，其他 4 个策略仍然通过影响行为策略实现控制，例如事前的权利限制会影响事中的用户使用行为实施，事后的责任内容会影响事中的平台管理行为实施。因此，划分事前、事

中和事后 3 个阶段并没有改变控制策略之间的基本关系，而是基于各策略的重点内容，旨在形成一个更加有序可循且集事前预防、事中管理与事后归责于一体的科学数据控制策略体系。

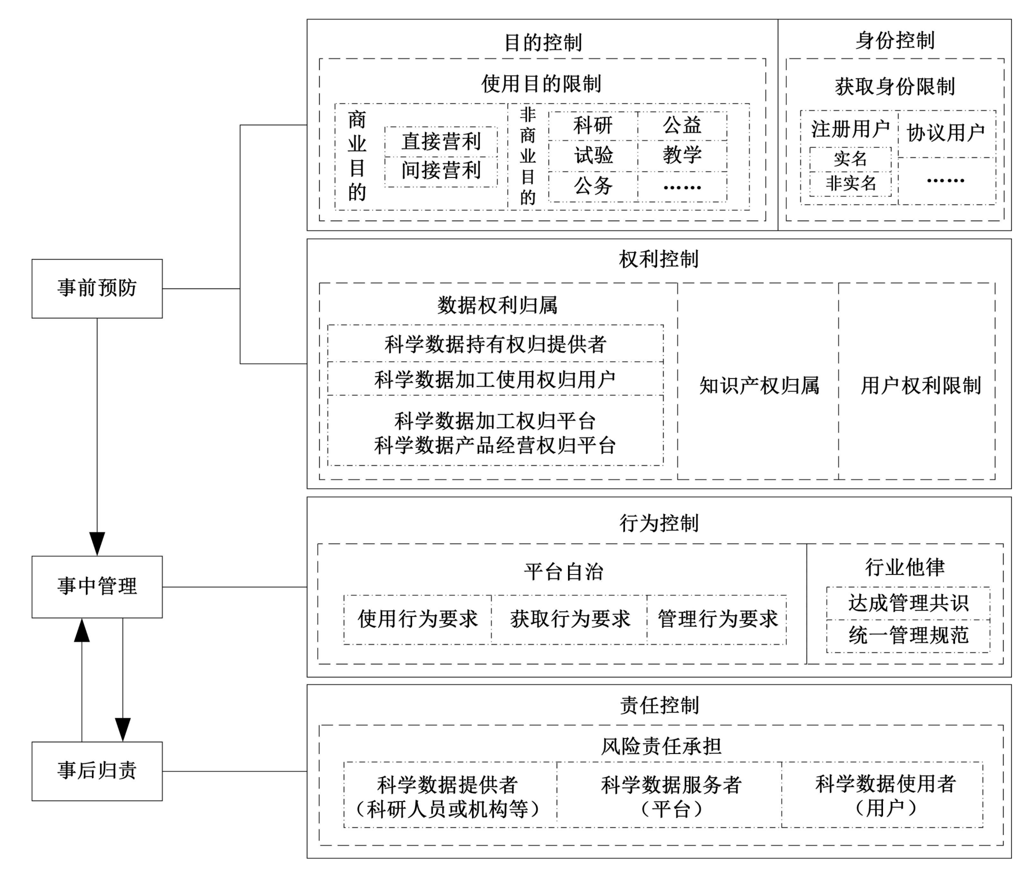


图 2 科学数据的“三阶段”控制策略模型

Figure 2 Model of the “three-stage” control strategy of scientific data

4.1 事前预防：细化三项策略内容

4.1.1 强化目的控制：概括列举目的类型

用户获取或使用科学数据的真实动机和目的内隐性强，强化目的控制的关键在于不能笼统地规定“禁止商用”，而应明确用户可用或不可用的具体目的类型，便于用户事先判断其基于特定目的可否获取或使用特定的科学数据。如果采用反向规定，仅简单规定“不得用于商业目的”，对于普通用户而言，判断是否为商业用途并非易事。目的是否属于商业用途，不能简单地以主体是否具有商业身份而“一言蔽之”。考虑到概括规定也需明确允许的使用目的，建议推行“概括+列举”模式列明目的类型。“概括”有利于提升解释空间，“列举”有利于类比判断。使用目的以是否具有营利性为判断标准，可分为直接营利与间接营利两种类型，商业目的包括但不限于直接或间接从事有偿分享、营销、交易等活动，可用目的包括不以营利为目的的科研、试验、教学、公务或公益等。

4.1.2 加强权利控制：明确数据权利归属

权利控制所涉及的是知识产权、数据的权利归属，因知识产权已经有法律规定，其归属可以依法确定，而数据权利因缺乏法律规定，其归属问题还处于探索实践阶段。平台所发布的科学数据并非当然归属于平台、用户或提供者，可以《数据二十条》[1]为政策依据，根据科学数据来源和生成特征，分别界定数据生产、流通、使用过程中提供者、平台和用户享有的合法权利，建立以持有、加工使用、产品经营等为内容的数据产权运行机制。暂不考虑提供者享有科学数据持有权的正当性，提供者不因科学数据上传平台而丧失持有权；赋予平台基于汇总的科学数据加工权，并对加工形成的科学数据集等产品享有经营权；允许用户基于约定目的，对从平台获取的数据享有加工使用权。如此形成科学数据的结构性分置产权制度，有利于激发不同主体的积极性。

4.1.3 补强身份控制：统一分级分类管理

将用户身份进行分类具有一定价值，如用户分类模型（TIS 模型）框架为在线用户创新社区（OUICs）的用户管理提供了理论指导，提高了社区用户的有效管理和利用，使其创新价值最大化[40]。身份控制背后体现的是平台分级分类管理科学数据的理念，设置不同用户访问权限对应不同等级的科学数据，有限度、有层次地实现科学数据公开或共享。尚未实施身份控制策略的平台，意味着其内部的科学数据分级分类与用户分级匹配还不完善。科学数据的内容涉及面广、专业程度高、荷载利益多元，且通常是跨学科、跨部门或跨行业的，这导致不同平台对于所控制科学数据的敏感性与风险性认知不一。因此，身份控制的完善路径应自下而上，由各平台先实施契合自身实际的科学数据分级分类方法，并延续已有的获取身份要求，区分注册用户、实名用户、协议用户等不同身份给予不同权限。补强身份控制的关键在于分级分类管理，有助于落实《数据二十条》中“把该管的管住、该放的放开”的要求[1]，进而形成统一的行业规范。

4.2 事中管理：优化行为策略结构

4.2.1 转型平台自治：平衡平台管理与用户控制

用户使用行为要求、用户获取行为要求和平台管理行为要求，虽然主体有差别，但是其共性点在于都是由国家级科学数据平台自行设定，可统称为平台自治。获取行为要求和使用行为要求的受控主体都是用户，二者共同构成了用户行为控制策略。当前，平台自治重心在于限制用户的获取、使用行为，而数据价值在于利用，国家也在积极推动数据开放共享，用户行为限制愈多，科学数据开放共享程度就难以相对提高。平台自治要坚持“把该管的管住、该放的放开”，在保障安全的前提下尽可能地“少约束、多开放”。一方面，平台承担着维护平台安全、保障科学数据安全等责任，这意味着平台自治仍然需控制用户行为；另一方面，科学数据中心是促进科学数据开放共享的重要载体，这意味着用户应该被更少地约束。故，平台自治应从“重用户行为控制，轻平台管理行为”转向“平台管理与用户控制相平衡”。

4.2.2 新增行业他律：促进平台意识与规范统一

从整体上来看，本次调查取样过程中，仅有 14 个国家级科学数据平台在官网中提供了用户协议，其他平台官网要么无法进入、要么无法在页面中找到相关文件。从个体上来看，平台行为控制策略呈现出“个体化”多于“统一化”的趋势。例如，有的平台规定了获取行为要求，有的平台对此则没有规定；又如，有的平台规定集中于某一项或某两项控制策略，有的平台规定包含了多项控制策略。即便各平台都偏重于行为控制，但却并未形成实施行为控制的共识，更没有发展出统一的行为控制规范。因此，可以考虑由科技部直属事业单位——国家科技基础条件平台中心牵头，在其“建设规范标准和管理方式”职责范围内，推动促进各平台形成统一行为管理共识，制定通用的行为管理规范。相对于前述平台自治来说，这种行业他律能确保各平台行为策略的最低要求。此外，允许各平台在通用行为管理规范的基础上增加“个体化”元素。

4.3 事后归责：改进责任策略分配

合理分配责任意味着主体均承担与其行为相应的责任，不排除己方责任，不将己方责任转移至他方。科学数据服务者（平台）作为科学数据提供者（科研人员或机构等）与数据使用者（用户）之间的桥梁，串联起了科学数据从产生、汇交到使用的整个过程。如前所述提供者、平台和用户应享有相应的数据权利，他们也应当承担相应的责任。责任分配是确保责任控制策略发挥作用的重要机制，应在提供者、平台和用户三方主体间合理分配责任。科学数据质量高低与提供者密切相关，故提供者应承担确保数据的真实性、有效性、格式规范性等责任；平台应以确保提供的科学数据可发现、可访问、可互操作、可重用为己任[41]，故基本责任应包括数据的审查责任、管理责任和安全保障责任等；用户则需要依法依规获取和使用科学数据，并严格遵守平台的相关规定，违反不得转让给第三方、使用需标明出处、不得随意修改数据等应承担责任。

5 结语

科学数据既不像个人数据那样涉及复杂的人格利益，又不像企业数据那样荷载丰富的财产利益，可能是科学数据的公益性使然，其关注度远不如个人数据或企业数据。然而，科学数据是一国科学发展的重要战略资源，科学数据平台承担着整合科学数据、提供科学数据服务的重要使命，是科学数据开放共享的重要载体[17]。我国已经出台了《科学数据管理办法》《国家科技资源共享服务平台管理办法》等政策文件，以及《科技计划形成的科学数据汇交技术与管理规范》等国家标准，在科学数据治理领域已形成了基本制度。科学数据控制策略的建立和完善是科学数据平台制度建设的核心，也是数据基础制度建设的关键内容。虽然学界关于科学数据的研究成果相当丰富，但对科学数据控制策略的研究稍显不足。

鉴于此，本文运用扎根理论方法，以国家级科学数据平台运行实践中的用户协议为样本，研究科学数据控制策略问题。可能的创新之处和后续研究空间归纳如下：①拓展了科学数据

研究主题的宽度，后续研究可推进科学数据控制策略的理论构建；②提炼出科学数据控制策略的目的、行为、权利、身份和责任 5 个构成要素，后续研究可在此基础上细化要素内容或提出新的要素；③总结“五控一体”控制策略模型，并在此基础上优化提出“三阶段”控制策略模型，后续研究修正或优化前述策略模型。不过本文仍有不足之处，研究样本仅为国家级科学数据中心与数据库，并未将上海研发公共服务平台、黑龙江省科技资源共享服务中心等地方平台纳入调查范畴。

参考文献：

[1] 中华人民共和国中央人民政府.中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见[EB\OL].[2023-10-01].https://www.gov.cn/zhengce/2022-12/19/content_5732695.htm.

(The Central People's Government of the People's Republic of China. Opinions of the Central Committee of the CPC and the State Council on building basic systems for data to give full play to the role of data resources[EB\OL].[2023-10-01].https://www.gov.cn/zhengce/2022-12/19/content_5732695.htm.)

[2] 中华人民共和国中央人民政府.国务院办公厅关于印发科学数据管理办法的通知[EB\OL].[2023-10-01].https://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.

(The Central People's Government of the People's Republic of China. The scientific data management rules[EB\OL].[2023-10-01].https://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.)

[3] 邓仲华,黄雅婷.“互联网+”环境下我国科学数据共享平台发展研究[J].情报理论与实践,2017,40(2):128-132.

(DENG Z H, HUANG Y T. Research on the development of scientific data sharing platform at home in the “Internet+” environment[J]. Information studies: theory & application,2017,40(2):128-132.)

[4] 张新兴.科学数据共享平台的建设与服务探讨[J].现代情报,2016,36(11):109-113.

(ZHANG X X. Research on development and service of scientific data sharing platform[J]. Journal of modern information,2016,36(11):109-113.)

[5] 李正超.国内科学数据共享平台建设现状及发展策略研究[J].图书馆理论与实践,2018(8):108-112.

(LI Z C. Study on the current situation and development strategy of domestic scientific data sharing platform[J]. Library theory and practice,2018(8):108-112.)

[6] 司莉,王雨娃.我国科学数据共享平台数据组织的现状及改进建议——基于国家科技基础条件平台的分析[J].图书馆建设,2018(10):52-58.

(SI L, WANG Y W. Current situation and suggestions of data organization on domestic scientific data sharing platforms——a study based on national science and technology infrastructure

platform[J]. Library development,2018(10):52-58.)

[7] 朱亮,孟宪学,赵瑞雪,等.国家农业科学数据共享中心资源建设探析[J].数字图书馆论坛,2017(11):15-20.

(ZHU L, MENG X X, ZHAO R X, et al. Research on resource construction for National Agricultural Science Data Sharing Center[J]. Digital library forum,2017(11):15-20.)

[8] 南卓铜,王亮绪,吴立宗,等.科学数据中心数据与知识集成[J].中国科技资源导刊,2010,42(5):15-21,36.

(NAN Z T, WANG L X, WU L Z, et al. Experiences in data and knowledge integration of the environmental and ecological science data center for West China[J]. China science & technology resources review,2010,42(5):15-21,36.)

[9] 李翼,吴丹.开放医学科学数据平台调查研究[J].图书情报工作,2015,59(18):24-29,50.

(LI Y, WU D. Research on open medical scientific data platform[J]. Library and information service,2015,59(18):24-29,50.)

[10] 孙叶华,刘桂锋,陈帅印.基于黄河流域专题数据的国家科学数据中心关联模型构建研究[J].数字图书馆论坛,2022(8):2-10.

(SUN Y H, LIU G F, CHEN S Y. Research on the construction of association model of national science data center based on the thematic data of Yellow River basin[J]. Digital library forum,2022(8):2-10.)

[11] 储节旺,杨婷婷.长三角跨区域科学数据中心建设研究[J].数字图书馆论坛,2022(6):58-64.

(CHU J W, YANG T T. Study on the construction of cross-regional scientific data center in Yangtze River Delta[J]. Digital library forum,2022(6):58-64.)

[12] 廖小罕,封志明,高星,等.野外科学观测研究台站（网络）和科学数据中心建设发展[J].地理学报,2020,75(12):2669-2683.

(LIAO X G, FENG Z M, GAO X, et al. The development of field scientific observation and research stations (networks) and scientific data centers[J]. Acta geographica sinica,2020,75(12):2669-2683.)

[13] 张计龙,殷沈琴,张用,等.社会科学数据的共享与服务——以复旦大学社会科学数据共享平台为例[J].大学图书馆学报,2015,33(1):74-79.

(ZHANG J L, YIN S Q, ZHANG Y, et al. Social scientific data sharing and serving——an example of Fudan university social scientific data platform[J]. Journal of academic libraries,2015,33(1):74-79.)

[14] 肖雪葵.关于湖南省科技信息与科学数据共享平台建设的研究报告[J].科技进步与对策,2007(2):124-127.

(XIAO X K. Research report on the construction of Hunan science and technology information and scientific data sharing platform[J]. Science & technology progress and policy,2007(2):124-127.)

[15] 司莉,李月婷,邢文明,等.我国科学数据共享平台绩效评估实证研究[J].图书馆理论与实践,2014(9):30-35.

(SI L, LI Y T, XING W M, et al. An empirical study on the performance evaluation of scientific data sharing platforms in China[J]. Library theory and practice,2014(9):30-35.)

[16] 司莉,华小琴.我国科学数据共享平台的服务效能分析[J].图书馆工作与研究,2014(4):24-26.

(SI L, HUA X Q. Analysis of the service effectiveness of national science and technology infrastructure platforms[J]. Library work and study,2014(4):24-26.)

[17] 王翠萍,宋雯琪,姜鑫妍,等.我国科学数据共享平台数据描述及引用规范调查研究[J].图书馆学研究,2022(11):76-85.

(WANG C P, SONG W Q, JIANG X Y, et al. Investigation and research on scientific data description and citation standards of domestic scientific data sharing platform[J].Research on library science,2022(11):76-85.)

[18] 刘小宇,纪雪梅,王芳.我国科学数据共享平台的被引用状况调查与分析——以国家科技基础条件平台为例[J].情报杂志,2020,39(9):168-175.

(LIU X Y, JI X M, WANG F. How does academic literature cite scientific data on the national science ＆technology infrastructure platform[J]. Journal of intelligence,2020,39(9):168-175.)

[19] 孙卫,罗之兰,张蔚.科学数据共享平台的数据管理研究[J].科学学与科学技术管理,2005(12):32-36.

(SUN W, LUO Z L, ZHANG W. Research on management of science data sharing platform[J]. Science of science and management of S.&T.,2005(12):32-36.)

[20] 文禹衡,金铭.科学数据中心用户注册协议的合规性评估研究[J].图书情报知识,2022,39(1):119-129.

(WEN Y H, JIN M. Compliance assessment of user registration agreements of scientific data centers[J]. Documentation, information & knowledge,2022,39(1):119-129.)

[21] 王亮绪,吴立宗,南卓铜,等.开源技术在地球科学数据中心中的应用[J].中国科技资源导刊,2010,42(3):17-23,35.

(WANG L X, WU L Z, NAN Z T, et al. Application of open source technologies in geoscientific data centers[J]. China science & technology resources review,2010,42(3):17-23,35.)

[22] 高峰,王剑.知识管理在项目管理中的应用研究——以农业科学数据共享中心项目为例[J].科技管理研究,2012,32(16):184-187.

(GAO F, WANG J. Study on the application of knowledge management in project management[J]. Science and technology management research,2012,32(16):184-187.)

[23] 刘敬仪,江洪,廖宇.德国地球科学领域科学数据中心调查与启示[J].数字图书馆论坛,2019(12):52-58.

(LIU J Y, JIANG H, LIAO Y. Research and enlightenment from Germany scientific data center for

the field of geoscience[J]. Digital library forum,2019(12):52-58.)

[24] 吴思竹,王安然,修晓蕾,等.欧美生物医学科学数据中心建设及启示[J].数字图书馆论坛,2022(4):2-10.

(WU S Z, WANG A R, XIU X L, et al. Construction and implications of biomedical scientific data centers in Europe and the United States[J]. Digital library forum,2022(4):2-10.)

[25] ARORA R. Data management: state-of-the-practice at open-science data centers[M]. Handbook on data centers. New York: Springer,2015:1095-1108.

[26] DOWNS R R. Implementing the group on earth observations data management principles: observations of a scientific data center[J]. The international archives of the photogrammetry, remote sensing and spatial information sciences,2017,XLII-3/W2:51-55.

[27] ARDESTANI S B, HAKANSSON C J, LAURE E, et al. B2SHARE: an open eScience data sharing platform[C]//IEEE 11th international conference on e-Science. Munich:IEEE,2015:448-453.

[28] VIRT&L-COMM. VHERLA: a virtual molecular science data center allocated on the GARR cloud[EB\OL].[2023-08-19].http://services.chm.unipg.it/ojs/index.php/virtlcomm/article/view/203.

[29] KIM C W, YOON H, JIN D, et al. Integrated management system for a large computing resources in a scientific data center[J]. The journal of super computing,2016,72(9):3511-3521.

[30] M CINQUE, D COTRONEO, F FRATTINI, et al. To cloudify or not to cloudify: the question for a scientific data center[J]. IEEE transactions on cloud computing,2016,4(1):90-103.

[31] CAROSI A, LUCARELLI F, ANTONELLI L A, et al. The very high energy source catalog at the ASI science data center[C]//Software and Cyberinfrastructure for Astronomy IV. Edinburgh, SPIE Digital Library,2016:781-785.

[32] 陈向明.扎根理论的思路和方法[J].教育研究与实验,1999(4):58-63,73.

(CHEN X M. Grounded theory: its train of thought and methods[J]. Educational research and experiment,1999(4):58-63,73.)

[33] 朱明,廖熙铸.高校图书馆管理制度内化的理论构建——基于扎根理论的探索性研究[J].图书情报工作,2019,63(5):32-41.

(ZHU M, LIAO X Z. Theoretical construction of internalization of university library management system: an exploratory study based on grounded theory[J]. Library and information service,2019,63(5):32-41.)

[34] 冯亚飞,胡昌平,仇蓉蓉.数字学术资源用户隐私关注影响因素模型构建——基于扎根理论[J].情报科学,2019,37(3):3-8.

(FENG Y F, HU C P, QIU R R. Model construction of user privacy concerns for digital academic resources: a grounded theory study[J]. Information science,2019,37(3):3-8.)

[35] 巫芯宇.基于移动终端的大学生深阅读行为影响因素研究——基于扎根理论的质性分析[J].图书情报工作,2021,65(24):80-88.

(WU X Y. Research on the influencing factors of college students' deep reading behavior based on mobile terminal: qualitative analysis based on grounded theory[J]. Library and information service,2021,65(24):80-88.)

[36] 文禹衡,贺亚峰.数据产权市场主体的认知调查与矫正[J].图书馆论坛,2022,42(3):31-42.

(WEN Y H, HE Y F. A survey of market entities' conception of data property rights and its correction[J]. Library tribune,2022,42(3):31-42.)

[37] 盛东方.我国图书情报研究中的扎根理论应用[J].图书馆论坛,2020,40(8):78-86.

(SHENG D F. The application of grounded theory in library and information science research in China[J]. Library tribune,2020,40(8):78-86.)

[38] 张敬伟,马东俊.扎根理论研究法与管理学研究[J].现代管理科学,2009(2):115-117.

(ZHANG J W, MA D J. Grounded theory research method and management research[J]. Modern management science,2009(2):115-117.)

[39] 文禹衡,付张祎.基于扎根理论的我国数据要素确权地方立法策略及其调整[J].图书情报工作,2023,67(7):53-66.

(WEN Y H, FU Z Y. Local legislation strategy and adjustment on the right of data elements in China based on grounded theory[J]. Library and information service,2023,67(7):53-66.)

[40] YANG J. A framework of user classification model of online user innovation communities based on user innovation value[J]. Open journal of social sciences,2020,8(5):232-244.

[41] WILKINSON M D, DUMONTIER M, AALBERSBERG I J, et al. The FAIR guiding principles for scientific data management and steward ship[J]. Scientific data,2016,3(1):1-9.

作者贡献说明：

文禹衡：提出研究选题和思路，撰写初稿及修改定稿；

徐坤杉：数据收集与分析，初稿撰写。

The Control Strategy of Scientific Data and Its Model Optimization

Wen Yuheng[1,2] Xu Kunshan[1]

[1] School of Intellectual Property, Xiangtan University, Xiangtan 411105

[2] Hunan Data Governance and Smart Judicial Research Center, Xiangtan 411105

Abstract : [Purpose/Significance] Proposing and optimizing the control strategy and model of scientific data will help to promote the basic system of scientific data platform and balance the protection and sharing of scientific data. **[Method/Process]** Taking the user agreements of the national scientific data center as samples, it obtained 76 initial concepts, 9 subcategories and 5

main categories, and then refined the control strategy model of scientific data. **[Result/Conclusion]** In China, scientific data platform has formed a "five-control-in-one" control strategy model with purpose, behavior, rights, identity and responsibility as the constituent elements, which can be optimized a "three-stage" control strategy model in the aspects of pre-defense, in-process supervision and post-event rules. Specifically, the measures include: summarizing the purposes types for purpose control; clarifying the ownership of data rights for rights control; unifying classification management for identity control; transforming platform autonomy by the balance of platform management and user control, and adding new industry heteronomy for the unity of platform awareness and norms; rationally allocating the responsibilities of platforms, users and scientific data providers.

Keywords: scientific data　control strategy　grounded theory behavior control　right control